COMP9444 Project Summary

Skin Lesion Classification Using Deep Learning

Hanwen Miao(z5495976) Yuxin Xi(z5542594) Yirou Li(z5563505)

Yi Fan(z5458765) Shuyuan Xing(z5507776)

I. Introduction

Melanoma is one of the most common skin cancers[1][2] in the world with a high fatality rate[3]. Early detection and diagnosis are crucial for effective treatment[4][5], but the traditional diagnoses of skin lesions are time-consuming and laborious. Therefore, we need to establish an automated classification system for skin lesions to assist doctors in identifying skin cancer, improving accuracy[6][7].

This project aims to develop a skin lesion classification model to distinguish skin cancer (such as melanoma[8]) from other types of benign lesions[9]. This project uses the ISIC dataset[10][11]. To balance the imbalance of categories, we merged similar categories into three categories, including melanoma, nevus, and benign lesions, to balance the label distribution and improve the generalization ability of the model through data augmentation.

In terms of model selection, we used different machine learning[12] and deep learning[13] methods, including SVM[14], random forest[15], traditional convolutional neural networks (CNN), and ResNet-50 architecture[16]. Experimental results showed that ResNet-50 had the best performance, with an accuracy of nearly 77%. It had high sensitivity and specificity and was much higher than other models.

In this project, we explored the application of deep learning in skin lesion image analysis and effectively improved the classification accuracy. Experimental results showed that deep networks can better capture skin features[17][18] and provide valuable implementation methods for medical image classification research.

II. Related Work

Skin lesion classification has gained significant attention due to its potential to aid early melanoma detection, a key factor in improving patient outcomes. Prior research has explored various machine learning(ML) and deep learning(DL) models to classify skin lesions, including melanoma, nevi, and benign types, using image data.

In 2017, Esteva et al. [19] introduced the first deep-learning CNN model (Inception v3) capable of matching the diagnostic performance of 21 board-certified dermatologists on malignant lesions. This model deconstructed digital images of skin lesions, used data augmentation through random rotation, and developed its diagnostic criteria during training.

Haenssle et al. (2018) evaluated a DL model for skin lesion detection[20], finding it outperformed most dermatologists in detecting common melanomas. Brinker et al. (2019) investigated transfer learning with ResNet50[21], showing increased accuracy on small datasets but noting high computational demands. Codella et al. (2019) studied class imbalance in similar lesion classification within the ISIC challenge[22], trying to improve generalization using Focal Loss or weighted loss functions.

These studies highlight DL models' effectiveness in skin lesion classification while noting limitations like data dependency, class imbalance, and computational needs. To improve generalization, future work could focus on diverse data sources, class-balancing methods, and enhancements to grid structures like ResNet50, making models more applicable in clinical diagnostics.

III. Methods

In the whole research process, it is a multi-classification task, four methods were used in skin lesion classification task, including two machine learning methods and two deep learning methods.

1. Machine learning method:

1) Support Vector Machine(SVM) [23]:

SVM uses kernel tricks to handle non-linearly separable data. SVM needs to extract data features manually. Therefore, it can’t guarantee the quality and precision of the data features.

2) Random forest [24]:

Random forest is composed of multiple independent decision trees, each of the decision trees is a weak classifier. All of the decision trees combined to form a strong classifier, that is random forest. It will select features randomly. It also needs to extract data features manually.

The machine learning method lacks the ability to automatically learn features, and capture spatial information. When dealing with large-scale data, methods must have stronger feature extraction capabilities and better generalization. Therefore, deep learning is a better choice than machine learning..

2. Deep learning method:

1) Tradition convolutional neural network(CNN)[25]:

Convolutional Neural Network (CNN) is a deep learning model designed for processing images and spatial data. CNN gradually extracts local and global features of data through convolutional layers, pooling layers, and fully connected layers, thereby effectively identifying complex patterns.

Traditional CNN makes it difficult to maintain stability as the number of network layers increases and is prone to the gradient vanishing problem, which limits the improvement in depth and accuracy.

2) ResNet50[26]:

ResNet-50 is a convolutional neural network consisting of 50 layers in depth. It uses the residual learning framework to solve the training problem of deep networks. By introducing residual blocks, ResNet-50 can effectively alleviate the gradient vanishing problem, allowing information to be efficiently transmitted in the network. This design allows the network to maintain high accuracy while deepening.

IV. Experimental Setup

1. Preprocessing

This paper works on ISIC Challenge Datasets(https://challenge.isic-archive.com/data/). This dataset has 10015 images as training datasets, with a resolution of 800 x 600 pixels. 193 images as validation datasets and 1512 images as testing datasets(fig. 1).

It has 7 labels for all the images, including MEL, NV, BCC, AKIEC, BKL, DF, and VASC. By observing the data distribution, the data set has been stratified sampling, but the data still has the problem of imbalanced distribution(fig. 2).

In order to process the data more conveniently, each similar label is grouped into one label, and finally, there are three labels. MEL has MEL, BCC, AKIEC. NV has NV. BENIGH has BKL, DF, VASC[27]. Then, to solve the imbalanced problem, do oversampling which calculates the number of samples for each category and generates category weight by the inverse of the category count to balance the label distribution. After over-sampling, the distribution of labels is balanced(fig. 3). In the last step of preprocessing, transform each picture by randomly adjusting the image's size, flip, rotation, brightness, contrast, and other attributes to generate different samples. In this way, it can improve the generalization and robustness of the model.

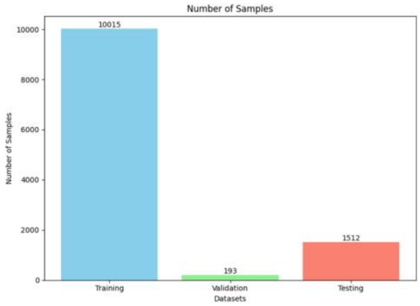


fig. 1 number of samples

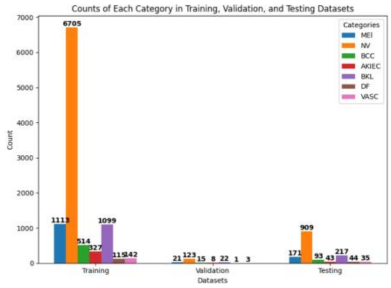


fig. 2 before over sampling

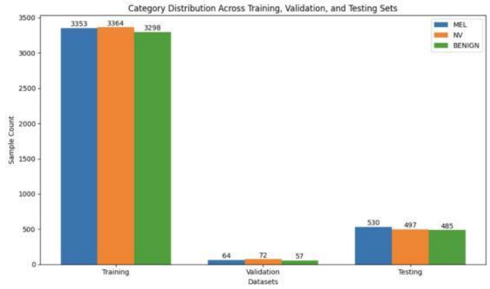


fig. 3 after over sampling

2. Traning:

1) Support Vector Machine(SVM):

Extracting features by flattening the image into a one-dimensional vector. Initialize the SVM model instance, specify the kernel function as a linear kernel, and set the penalty parameter C is 1.0 and kernel parameters is 0.01. Then train the model. Use cross-validation to optimize hyperparameters.

2) Random forest:

Extracting features by flattening the image into a one-dimensional vector. Initialize the Random forest classifier, set the hyperparameter number of trees is 100 and the number of minimal samples in each node is 2.

3) Tradition convolutional neural network(CNN):

Use four convolutional layers:(kernel size and stride of each max pooling layer is)

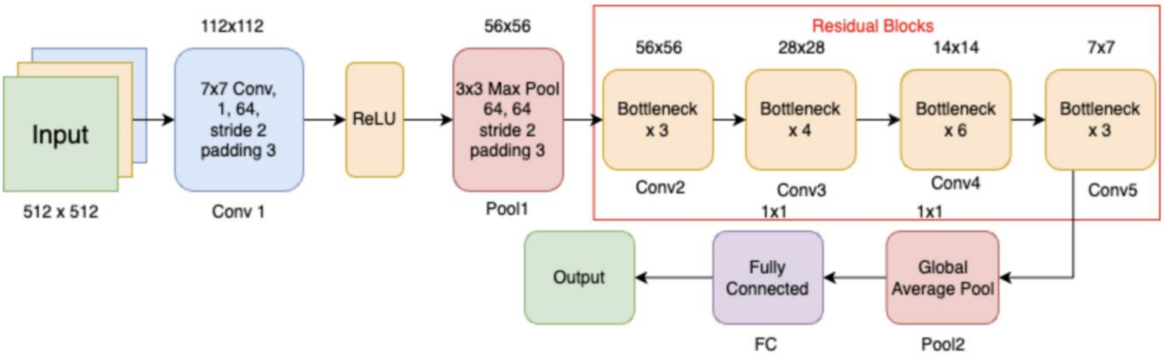
Layers	in_channels	out_channels	kernel_size	stride	padding
Con1	3	64	3	1	1
Con2	64	128	3	1	1
Con3	128	256	3	1	1
Con4	256	512	3	1	1

Use adam optimizer with 0.001 learning rate and CE loss function.

4) ResNet50:

Use ResNet50 as the backbone which has 50 Convolutional layers, one batch normalization layer, four residual module layers, and one global average pooling layer with a rule activation function. One linear link layer with three outputs, including MEl, NV, and BENIGN. Use Adam optimizer with 0.0001 learning rate and BCE loss function. Batch size is equal to 32. To prevent overfitting, set an early stopping mechanism with 9 degrees of tolerance.

The structure of ResNet-50 is shown as below:

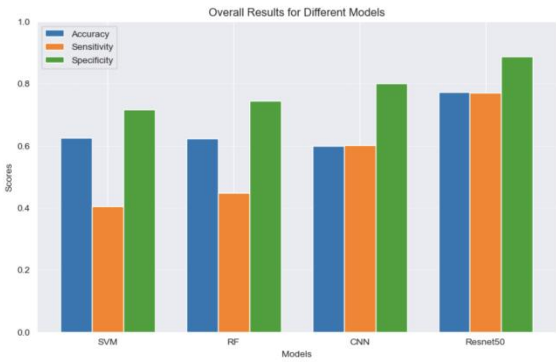


For evaluation strategy, use accuracy, sensitivity, and precision[28]. Accuracy represents the overall accuracy of the model's predictions, Sensitivity represents the model's ability to identify all positive samples, and Precision represents the proportion of true positive samples in the positive samples predicted by the model.

V. Results

1) 4 Models performance on test set

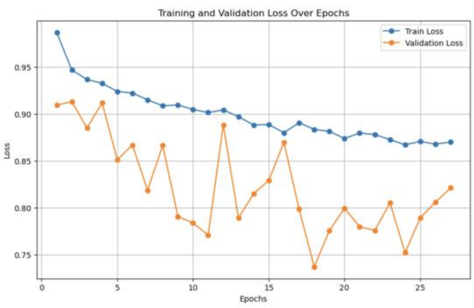
Metric	SVM	RF	CNN	ResNet-50
Accuracy	0.6243	0.6224	0.5979	0.7718
Sensitivity	0.4049	0.4482	0.6006	0.7704
Specificity	0.7161	0.7444	0.7995	0.8861



ResNet-50 excels in sensitivity, specificity, and accuracy, with a notable sensitivity of approximately 77%. CNN performs moderately, achieving a sensitivity of 60%, which is significantly better than Random Forest and SVM (around 4%). The low sensitivity of Random Forest and SVM indicates a tendency to misclassify malignant samples as non-malignant, underscoring their limitations in handling this critical category; ResNet-50 again leads in specificity, showcasing its balanced performance across different types of lesions. The accuracy of ResNet-50 is higher than the other three methods. That shows the ResNet-50 model has the best performance in these four methods.

2) Loss for CNN and RestNet50 during epochs:

Improved Generalization: In the CNN model, the validation loss plateaus after around 10 epochs, while the training loss continues to decline, indicating mild overfitting. In contrast, ResNet-50 shows closely aligned training and validation loss curves, which remain stable over the epochs. This alignment demonstrates ResNet-50's stronger generalization capability on unseen data.



CNN

Resnet50

Higher Accuracy in Classification: The ResNet-50 model achieves greater sensitivity and specificity, especially in detecting malignant cases, due to its deeper layers and ability to learn refined features. The stable validation loss indicates that ResNet-50 can differentiate between subtle class differences more reliably than the CNN model.

Reduced Overfitting: The continuous drop in training loss and relatively flat validation loss in the CNN reveals a tendency to memorize the training data. ResNet-50 with residual connections that prevent information loss, avoids this problem, resulting in a model that learns general patterns rather than overfitting to the training set.

VI. Conclusions

ResNet-50 outperforms other methods in overall accuracy (0.7718), sensitivity (0.7704), and specificity (0.8861), due to its deep convolutional architecture that captures complex image features. This makes it suitable for skin disease classification, where subtle differences are key. In contrast, traditional machine learning methods like SVM and RF, although efficient in low-resource environments, rely on handcrafted features and show higher specificity (0.7161 for SVM, 0.7444 for RF) but lower sensitivity (0.4049 for SVM, 0.4482 for RF), limiting their detection accuracy for disease cases.

The CNN model also shows relatively high specificity (0.7995) but is outperformed by ResNet-50 in sensitivity and accuracy, likely due to its shallower architecture. However, deep learning models like ResNet-50 have high computational demands and data dependencies, requiring further data augmentation to improve performance on imbalanced datasets.

Future work could involve diverse data augmentation, ensemble methods combining multiple model predictions, and fine-tuning techniques. Deployment optimization strategies, such as model quantization and pruning, could enhance ResNet-50's feasibility for resource-limited settings. Overall, while ResNet-50 demonstrates strong performance, further optimizations can make it more adaptable and efficient for practical applications.

References

1. Gloster Jr, Hugh M., and Kenneth Neal. "Skin cancer in skin of color." Journal of the American Academy of Dermatology 55.5 (2006): 741-760.

2. Ashraf, R.; Afzal, S.; Rehman, A.U.; Gul, S.; Baber, J.; Bakhtyar, M.; Mehmood, I.; Song, O.Y.; Maqsood, M. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. IEEE Access 2020, 8, 147858–147871.

3. Codella, Noel, et al. "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)." arXiv preprint arXiv:1902.03368 (2019).

4. Greinert, Rüdiger. "Skin cancer: new markers for better prevention." Pathobiology 76.2 (2009): 64-81.

5. Janda, Monika, et al. "Early detection of skin cancer in Australia–current approaches and new opportunities." Public health research & practice 32.1 (2022).

6. Kassem, Mohamed A., et al. "Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review." Diagnostics 11.8 (2021): 1390.

7. Oliveira, Roberta B., et al. "Computational methods for pigmented skin lesion classification in images: review and future trends." Neural Computing and Applications 29 (2018): 613-636.

8. Gouda, Walaa, et al. "Detection of skin cancer based on skin lesion images using deep learning." Healthcare. Vol. 10. No. 7. MDPI, 2022.

9. Adegun, Adekanmi, and Serestina Viriri. "Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art." Artificial Intelligence Review 54.2 (2021): 811-841.

10. Kassem, Mohamed A., Khalid M. Hosny, and Mohamed M. Fouad. "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning." IEEE Access 8 (2020): 114822-114832.

11. Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.0336

12. Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349.6245 (2015): 255-260.

13. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

14. Joachims, Thorsten. Making large-scale SVM learning practical. No. 1998, 28. Technical report, 1998.

15. Biau, Gérard, and Erwan Scornet. "A random forest guided tour." Test 25 (2016): 197-227.

16. Al-Haija, Qasem Abu, and Adeola Adebanjo. "Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network." 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). IEEE, 2020.

17. Benyahia, Samia, Boudjelal Meftah, and Olivier Lézoray. "Multi-features extraction based on deep learning for skin lesion classification." Tissue and Cell 74 (2022): 101701.

18. Yu, Lequan, et al. "Automated melanoma recognition in dermoscopy images via very deep residual networks." IEEE transactions on medical imaging 36.4 (2016): 994-1004.

19. Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017). https://doi.org/10.1038/nature21056

20. HA Haenssle, C Fink, R Schneiderbauer, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol, 29 (2018), pp. 1836-1842

21. Brinker, T. J., Hekler, A., Enk, A. H., et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task." European Journal of Cancer, 2019.

22. Codella, N., Rotemberg, V., Tschandl, P., et al. "Skin lesion analysis toward melanoma detection: A challenge hosted by the international skin imaging collaboration (ISIC)." 2018.

23. Burges, C.J. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998).https://doi.org/10.1023/A:1009715923555

24. Breiman,L.Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:10109

33404324

25. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

26. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

27. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161 (2018).

28. Lopez, Adria Romero, et al. "Skin lesion classification from dermoscopic images using deep learning techniques." 2017 13th IASTED international conference on biomedical engineering (BioMed). IEEE, 2017.